AD-A065 681 MASSACHUSETTS INST OF TECH CAMBRIDGE LAB FOR INFORMA--ETC F/G 9/2
NEW CONCEPTS IN NONLINEAR INFINITE-HORIZON STOCHASTIC ESTIMATIO--ETC(U)
1978 L K PLATZMAN AFOSR-77-3281

UNCLASSIFIED LIDS-P-868 AFOSR-TR-79-0080 NL

| OF |
ADA
065681

END
DATE
FILMED
4-79
DDC

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: AFOSR-TR-79-0080	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): NEW CONCEPTS IN NONLINEAR INFINITE-HORIZON STOCHASTIC ESTIMATION AND CONTROL: THE FINITE ELEMENT CASE	5. TYPE OF REPORT & PERIOD COVERED: Interim rept.
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Loren K. Platzman	8. CONTRACT OR GRANT NUMBER(s): AFOSR-77-3281, AFOSR-72-2273
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Massachusetts Institute of Technology, Laboratory for Information & Decision Systems, Cambridge, MA 02139	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 61102F, 2304, A1
11. CONTROLLING OFFICE NAME AND ADDRESS: Air Force Office of Scientific Research/NM, Bolling AFB, Washington, DC 20332	12. REPORT DATE: 1978
	13. NUMBER OF PAGES: 7
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): LIDS-P-868, 29 p.	15. SECURITY CLASS. (of this report): UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

This paper will appear in the Proceedings of the IEEE Conference on Decision and Control, San Diego, CA, Jan. 1979

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A finite probabilistic system (FPS) is a discrete-time controlled stochastic process having finite input, output, and (internal) state sets. (A partially-observed Markov decision process is an example of an FPS). It may be viewed as the simplest formulation of a nonlinear estimation and control problem.

Under conditions similar to observability and controllability in linear systems, the problem of selecting inputs, on the basis of past inputs and outputs (with perfect recall), so as to maximize a time-averaged expected reward, is shown to be meaningful as the horizon increases without bound or as a discount approaches unity: an optimal

DD FORM 1 JAN 73 1473

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

420 950

20. Abstract continued.

strategy exists; it may be realized by a (strategy-independent) state estimator along with a stationary policy on the state estimate; and its performance does not depend on the initial state of information.

Dual control aspects of the problem, and potential extention of the results to more general systems are briefly discussed.

LIDS-P-868

(Proc. IEEE Conference on Decision and Control, San Diego, CA, Jan. '79)

NEW CONCEPTS IN NONLINEAR INFINITE-HORIZON STOCHASTIC ESTIMATION AND CONTROL: THE FINITE ELEMENT CASE

Loren K. Platzman

Bell Telephone Laboratories
Naperville, IL 60540

Department of Industrial and
Operations Engineering
The University of Michigan
Ann Arbor, MI 48109

Abstract

A finite probabilistic system (FPS) is a discrete-time controlled stochastic process having finite input, output, and (internal) state sets. (A partially-observed Markov decision process is an example of an FPS). It may be viewed as the simplest formulation of a nonlinear estimation and control problem.

Under conditions similar to observability and controllability in linear systems, the problem of selecting inputs, on the basis of past inputs and outputs (with perfect recall), so as to maximize a time-averaged expected reward, is shown to be meaningful as the horizon increases without bound or as a discount approaches unity: an optimal strategy exists; it may be realized by a (strategy-independent) state estimator along with a stationary policy on the state estimate; and its performance does not depend on the initial state of information.

Dual control aspects of the problem, and potential extention of the results to more general systems are briefly discussed.

I. INTRODUCTION

The deceptive simplicity of the linear-quadratic-Gaussian problem formulation and solution has been articulated by Witsenhausen [18], among others. This paper describes recent work (much of which was originally reported in the author's doctoral dissertation [11]) aimed at understanding the relationship between estimation and control in a more general setting. Specifically, it examines a class of discrete-time undiscounted infinite-horizon stochastic control problems in which the input, output, and state sets are all finite. Conditions similar to controllability and observability are introduced and shown to imply well-posedness of the problem in the following sense: The optimal performance converges to that of a stationary policy on the sufficient statistic, as the horizon grows without bound or the discount approaches unity.

This approach clarifies the concept of "dual control" [7, 3, 17] in undiscounted infinite-horizon stochastic control. Any "dual control" problem can be slightly modified so that the conditions described above are satisfied. On the other hand, some unmodified "dual control" problems are meaningless unless a finite horizon or discount rate is specified.

Consider, for example, a fair coin that is tossed at times k=0,1, ... The outcome of toss k is denoted s(k)=H or T. Immediately after toss k>0, an experimenter observes y(k) where

$$y(k) = \left\{ \begin{array}{l} 0, \text{ if } s(k-1)=s(k) \\ 1, \text{ if } s(k-1)\neq s(k) \end{array} \right\} .$$

The experimenter then selects an input from the set {H,T,B}. The object is to maximize the limiting frequency of correct guesses u(k)=s(k). State information is gained by selecting u(k)=B, which causes a biased coin (e.g. Pr{s(k+1) is H}=.6) to be used in toss k+1.

If the horizon is finite or a discount β is used, then the problem is well-posed; the biased coin is used during a finite interval, and the most likely state is selected thereafter. As the horizon grows without bound or $\beta \uparrow 1$, the limiting strategy becomes: u(k)=B, indefinitely. Since there will be no guesses, and hence no correct guesses, this is the worst possible strategy.

An optimal strategy is:

$$u(k) = \left\{ \begin{array}{ll} \text{B}, & \text{if k is a power of 2} \\ \text{the most likely state}, & \text{otherwise} \end{array} \right\} .$$

The limiting proportion of correct guesses is now 1. This strategy suffers the aesthetic drawback of being nonstationary. And it clearly is not approached as the horizon grows without bound or the discount approaches unity. For these reasons, the problem is considered to be ill-posed in the conventional undiscounted infinite horizon formulation.

The problem becomes more tractible if we add to the plant model a mechanism whereby observa-

Research supported by AFOSR grant 77-3281

tion dynamics fail (in a specifically described manner, e.g. equally likely observation of 0 or 1) with probability ε, with $0 < \varepsilon << 1$. This version of the problem has a solution that agrees with LQG-induced intuition. The optimal strategy is stationary, and alternates between measurement and guessing with an average period that grows without bound as $\varepsilon \to 0$. Because the system is fallible, the mathematics of optimization will not reach into the arbitrarily distant past for information that in practice would surely have become noise-corrupted.

This paper will describe conditions that imply desirable structural properties of the type discussed above. Results are stated without proof; for details, see [11,12,13]. Our presentation follows a standard plan:

- Problem Formulation: Give the plant model and performance criterion.
- State Estimation: Derive a recursive form for the sufficient statistic and specify a condition for stability of the state estimation process.
- Dynamic Programming Formulation: Define an operator whose fixed point is the solution to the infinite horizon problem.
- Fixed Point Theorem: Prove that the dynamic programming operator has a unique fixed point.
- Computational Considerations: Show how an ε-optimal solution can be obtained on a digital computer.

II. PROBLEM FORMULATION

a) The plant model

(2.1) Definition. A finite probabilistic (dynamical) system (FPS) is a 4-tuple $(U, Y, S, \{P(y|u) : y \in Y, u \in U\})$ where:

(i) U is a finite nonempty set of input values (or decisions);

(ii) Y is a finite nonempty set of output values (or observations);

(iii) $S = \{1, \ldots, N\}$ is a finite nonempty set of (internal) state values;

(iv) Each $P(y|u)$ is an $N \times N$ substochastic matrix of state transition probabilities, and

$$(2.2) \qquad p(u) = \Sigma_{y \in Y} P(y|u)$$

is a stochastic matrix, $\forall u \in U$.

The dynamic evolution of an FPS is described in the following terminology:

1. When a decision-maker specifies input $u(k)$, that input is said to be accepted by the FPS. Output $y(k+1)$ is subsequently emitted by the FPS.

2. Given that an FPS in state $s(k)=i$ accepts input $u(k)=u$, it will undergo a transition to state $s(k+1)=j$ and emit output $y(k+1)=y$, with (conditional) probability $P_{ij}(y|u)$, (conditionally) independently of the "past" $\{s(k'), u(k'), y(k'+1)\}_{k'=0}^{k-1}$.

3. The Markov decision process (MDP) consisting of the internal state and input processes of an FPS is called the underlying process (of that FPS). It is described by the stochastic matrices $\{P(u): u \in U\}$.

4. The time set is $\{0, \ldots, K\}$. The terminal time K is called the horizon.

Remark. This notation is due to Paz [10].

b) The probability spaces

An FPS is studied in conjunction with an initial state probability (ISP) and a control strategy (CS).

The ISP, denoted by π, is a stochastic N-vector having the interpretation $\pi_i = \Pr\{s(0)=i\}$. The set of ISP's (i.e. the set of horizontal stochastic N-vectors) is denoted by Π.

The CS, denoted by γ, is a mapping $\gamma: Z^* \to U$, where Z^* represents the free monoid generated by $U \times Y$, i.e. the set of finite strings of I/O pairs. A decision-maker acting according to γ selects inputs

$$(2.3) \qquad u(k) = \gamma[z(k)]$$

where $z(k)$ is the information vector

$$(2.4) \qquad z(k) = (u(0),y(1))\ (u(1),y(2))\ \cdots\ (u(k-1),y(k)).$$

The set of CS's (i.e. the set of mappings from Z^* to U) is denoted by Γ.

We may view $\{s(k),u(k),y(k)\}$ as random variables on a probability space $\underline{P}[\pi,\gamma] = (\Omega, F, \Pr_{\pi,\gamma})$ where: Ω is the infinite product set of $S \times U \times Y$; F is the σ-algebra generated by the finite cylinders; and $\Pr_{\pi,\gamma}$ is determined in a straightforward manner from the transition probabilities described above.

$E_{\pi,\gamma}$ will denote the expectation operator associated with $\Pr_{\pi,\gamma}$.

c) The performance indices

Consider a bounded real-valued function R on $S \times U \times Y \times S$, and define

(2.5) $r(k) = R[s(k), u(k), y(k+1), s(k+1)]$

(2.6) $g(K) = K^{-1} \sum_{k=0}^{K-1} r(k)$

(2.7) $\tilde{g}(\beta) = (1-\beta)^{-1} \sum_{k=0}^{\infty} \beta^k r(k)$ $\qquad \beta<1$

We call r(k) an incremental reward; g(K) is the time-averaged reward, and $\tilde{g}(\beta)$ is the discount-averaged reward. Each is a random variable on $\underline{P}[\pi,\gamma]$.

d) Statement of the problem

The problem is to demonstrate the existence of strategies that "optimize" the infinite-horizon performance indices $\lim_{K\to\infty} g(K)$ and $\lim_{\beta\uparrow 1} \tilde{g}(\beta)$. Specifically, we determine conditions that assure the existence of an optimal performance g, and a family $\{\gamma^\pi\}$ of optimal CS's such that, for all ISP's π and all CS's γ,

(2.8) $\lim_{K\to\infty} E_{\pi,\gamma^\pi}\{g(K)\} = \lim_{\beta\uparrow 1} E_{\pi,\gamma^\pi}\{\tilde{g}(\beta)\} = g$

(2.9) $\limsup_{K\to\infty} E_{\pi,\gamma}\{g(K)\} \le g$

(2.10) $\limsup_{\beta\uparrow 1} E_{\pi,\gamma}\{\tilde{g}(\beta)\} \le g.$

e) Bibliographic notes

Standard references on the role of MDP's in stochastic control theory are Bertsekas [4] and Kushner [9]. The Partially Observed MDP was independently conceived by Drake [6] and Astrom [1,2]. Computational algorithms that solve finite-horizon and discounted POMDP's have been given by Smallwood and Sondik [15] and Sondik [16]. A more extensive bibliography may be found in [11, 12,13].

III. THE STATE ESTIMATOR FOR FPS's

a) The recursive formula

Let us introduce some terminology:

(3.1) For $z=(u_1,y_1)(u_2,y_2)\cdots(u_k,y_k)\in Z^*$, define the matrix product $P(z)=P(y_1|u_1)\cdot P(y_2|u_2)\cdot\ldots\cdot P(y_k|u_k)$.

(3.2) Define the vertical N-vector $\nu = (1, \ldots, 1)^t$.

(3.3) Define $T(\pi,z) = \pi P(z) \,/\, \pi P(z)\nu$, when $\pi P(z) \neq 0$.

(3.4) Define random variables on $\underline{P}[\pi,\gamma]$:

$$\eta^\pi(k) = T(\pi, z(k))$$

Now $\eta^\pi(k)$ is the vector of conditional state probabilities at time k, given inputs and outputs that have evolved up to that time. It may be computed by the (strategy-independent) recursive formula

(3.5) $\eta^\pi(k) =$

$$\begin{cases} \pi, & \text{if } k=0 \\ T(\eta^\pi(k-1), (u(k-1),y(k))), & \text{otherwise} \end{cases}.$$

b) A metric on Π

(3.6) Definition (Bayes' operator). For $\pi\in\Pi$, $w\in R_N$, with $w_i \ge 0$ $\forall i\in S$ and $\pi w>0$, let $\pi\circ w$ denote the vector in Π having entries

$$(\pi\circ w)_i = \pi_i w_i / \pi w.$$

(3.7) Definition. For $\pi,\pi'\in\Pi$, define

(a) $|\pi - \pi'| = \sum_{i\in S} |\pi_i - \pi'_i|$;

(b) $\delta[\pi,\pi'] = \sum_{i\in S} \max(\pi_i - \pi'_i, 0)$

(c) $\Delta[\pi,\pi'] = \sup\{\delta[\pi\circ w, \pi'\circ w]:$

$$w\in R_N,\ w_i\ge 0\ \forall i\in S,\ \pi w>0,\ \pi' w>0\}.$$

(3.8) Lemma. $|\cdot - \cdot|$, δ and Δ are metrics on Π, and

$$0 \le \tfrac{1}{2}|\pi - \pi'| = \delta[\pi,\pi'] \le \Delta[\pi,\pi'] \le 1.$$

(3.9) Theorem (evaluation of Δ). For $\pi,\pi'\in\Pi$, define:

$$c_1 = \min\{\pi'_i/\pi_i : \pi_i>0\};$$
$$c_2 = \min\{\pi_i/\pi'_i : \pi'_i>0\}.$$

Then

$$\Delta[\pi,\pi'] = \frac{1 - \sqrt{c_1 c_2}}{1 + \sqrt{c_1 c_2}}.$$

The metric δ, also known as the Hajnal measure, has many applications in the theory of ergodic Markov chains [8]. Informally, $\delta[\pi,\pi']$

is the (minimal) "quantity of probability" that would have to be "reassigned" in order to transform probability distribution π into probability distribution π'. Similarly, $\Delta[\pi,\pi']$ is the least upper bound on the quan tity of conditional probability by which π and π' might differ if they were conditioned on identical observations.

The distinction between δ and Δ is also illuminated by an examination of the topologies they induce on Π: the topology induced by δ is connected, but Δ causes Π to be separated into 2^N-1 "faces".

c) The contraction property of T

It is well known that if P is a stochastic matrix and

$$(3.10)\quad \hat{a}[P] = \max_{i,j\in S} \delta[e^iP,\ e^jP] < 1$$

then, for any $\pi,\pi'\in\Pi$,

$$(3.11)\quad \delta[\pi P,\ \pi'P] \le \hat{a}[P]\ \delta[\pi,\pi'],$$

i.e., the transformation $f[\pi] = \pi P$ is a contraction in Π. One consequence of this property of P is that $\{\pi(P)^n\}$ approaches a unique limit as $n\to\infty$. The rate of covergence $\hat{a}[P]$ is called the ergodic coefficient of the stochastic matrix P.

(3.12) Definition: If P is a nonzero substochastic matrix, then define

$$\alpha[P] = \max\{\Delta[T(e^i,P),\ T(e^j,P)] : e^iP\neq 0,\ e^jP\neq 0\}$$

Remark: The evaluation of $\alpha[P]$ by (3.9) requires N^3 operations. This is comparable to the effort expended when multiplying two N×N matrices.

The generalized ergodic coefficient $\alpha[P]$ has the following properties:

(3.13) Lemma. (a) $0 \le \alpha[P] \le 1$ for all substochastic matrices $P\neq 0$.

(b) $\alpha[P] < 1 \iff$ P is subrectangular*.

(c) $\alpha[P] = 0 \iff \text{rank}[P] = 1$.

(3.14) Theorem. (Contraction Property of T)

$$\Delta[T(\eta,P),\ T(\eta',P)] \le \alpha[P]\ \Delta[\eta,\eta'],\quad \eta P\neq 0,\ \eta'P\neq 0.$$

*In a subrectangular matrix, $P_{ij}>0$ and $P_{mn}>0$ imply $P_{in}>0$ and $P_{jm}>0$.

(3.15) Corollary.

$$\alpha[P] = \sup\{\Delta[T(\eta,P),\ T(\eta',P)] : \eta P\neq 0,\ \eta'P\neq 0\}.$$

(3.16) Corollary. $\alpha[PQ] \le \alpha[P]\ \alpha[Q]$.

d) Another metric on Π

With c_1, c_2 as in (3.9), define

$$(3.17)\quad D[\pi,\pi'] = 1 - \min[c_1,c_2].$$

Now D is a metric on Π and $(1/4)\ D[\pi,\pi'] \le \Delta[\pi,\pi'] \le D[\pi,\pi'] \le 1$. It has the following remarkable property (required in Theorem (5.2)): If v is a convex function on Π and $|v| = \sup_{\pi,\pi'\in\Pi}\{v(\pi)-v(\pi')\}$ then $|v(\pi)-v(\pi')| \le |v| \cdot D[\pi,\pi']$. This occurs because the *discontinuities** of Δ (discussed in section 3b) coincide with the potential discontinuities of a convex function on Π.

e) The condition on observation dynamics

As in (3.12) define

$$(3.18)\quad a[P] = \max\{D[T(e^i,P),\ T(e^j,P)] : e^iP\neq 0,\ e^jP\neq 0\}$$

Now consider the following condition

(3.19) Condition (detectability). There is an $a<1$ and an integer ζ such that, for every ISP π and every CS γ:

$$E_{\pi,\gamma}\{a[P(z(\zeta))]\} \le a.$$

Assuming (3.19) holds, there exists an $\alpha < a$ such that

$$(3.20)\quad E_{\pi,\gamma}\{\alpha[P(z(\zeta))]\} \le \alpha.$$

Using the recursion (3.5) and the contraction (3.14), we obtain

$$(3.21)\quad \lim_{k\to\infty} E_{\pi,\gamma}\{|\eta^\pi(k)-\eta^{\pi'}(k)|\} = 0 \quad \forall\ \pi,\pi'\in\Pi,\ \gamma\in\Gamma.$$

This is analogous to convergence of the conditional state distribution (and not simply the conditional mean) to an initial-value-independent

*with respect to conventional metrics on Π.

trajectory in the Kalman filter.

An FPS may be trivially modified so that Condition (3.19) is satisfied. For $0 < \varepsilon << 1$, multiply each $P(y|u)$ by $1-\varepsilon$ and then add $\varepsilon/(\#S \cdot \#Y)$ to each entry of each $P(y|u)$. This quantity may be interpreted as the probability of model failure, as discussed in Section I.

IV. DYNAMIC PROGRAMMING FORMULATION

Define:

(4.1) e^i is the "unit vector" in Π whose i-th entry equals unity.

(4.2) V is the vector space of real-valued bounded continuous functions on Π.

(4.3) $\bar{V} = \{v \in V : v(e^N)=0\} \subset V$.

(4.4) $\theta \in V$ is the "zero function" $\theta(\pi)=0$, $\forall \pi \in \Pi$.

(4.5) $q(u)$ is the expected incremental reward vector, a vertical N-vector with entries

$$q_i(u) = \Sigma_{j \in S} \Sigma_{y \in Y} P_{ij}(y|u)\, R(i,u,y,j).$$

(4.6) $\tilde{f}_\beta : V \to V$ is the discounted dynamic programming operator

$$[\tilde{f}_\beta v]\,(\pi) = \max_{u \in U} \{\pi q(u) + \beta \Sigma_{y \in Y} (\pi P(y|u) \vee)\, v(T(\pi,(u,y)))\}.$$

(4.7) $f : V \to V$ is the undiscounted dynamic programming operator, given by $f = \tilde{f}_1$.

(4.8) $\bar{f} : V \to \bar{V}$ is the normalized (undiscounted) dynamic programming operator given by

$$[\bar{f}v]\,(\pi) = [fv]\,(\pi) - [fv]\,(e^N).$$

Remark: This operator corresponds to a value-iteration algorithm of D. J. White [19].

(4.9) $\bar{f}_\lambda : \bar{V} \to \bar{V}$ is the damped normalized (undiscounted) dynamic programming operator given by

$$\bar{f}_\lambda \bar{v} = \lambda\, \bar{f}\, \bar{v} + (1-\lambda)\, \bar{v}$$

Remark: This operator corresponds to a value-iteration algorithm of P. J. Schweitzer [14].

Following Astrom (1966),

(4.10) $G_K(\pi) = \max_\gamma E_{\pi,\gamma}\{g(K)\} = K^{-1}\,[f^K\theta]\,(\pi)$.

Similarly, using the contraction property of discounted dynamic programming operators, we see that $\tilde{f}_\beta$ has a unique fixed point $\tilde{v}^*_\beta$, satisfying

(4.11) $\tilde{v}^*_\beta = \lim_{K \to \infty} (\tilde{f}_\beta)^K v \qquad \forall\, v \in V$

and

(4.12) $\tilde{G}_\beta(\pi) = \max_\gamma E_{\pi,\gamma}\{\tilde{g}(\beta)\} = (1-\beta)\, \tilde{v}^*_\beta(\pi)$.

This last equation is justified as outlined in Chapter 6 of [4].

Both G_K and $\tilde{G}_\beta$ are known to be convex and continuous on Π.

V. THE FIXED POINT THEOREM

We now require a second condition:

(5.1) Condition (reachability). There is a $\rho<1$ and an integer ξ such that, for every $\pi \in \Pi$, $j \in S$, a sequence of inputs $u_1, \ldots, u_\xi$ exists, satisfying

$$1 - \Sigma_{i \in S}\, \pi_i [P(u_1) \cdot \ldots \cdot P(u_\xi)]_{ij} \le \rho.$$

Also define

$$Q_{max} = \max_{i \in S} \max_{u \in U} q_i(u) \qquad Q = Q_{max} - Q_{min}$$

$$Q_{min} = \min_{i \in S} \min_{u \in U} q_i(u) \qquad C = \frac{(\xi+\tau)\,Q}{(1-\rho)(1-a)}.$$

The following theorem is the main result of this research.

(5.2) Theorem. Assume Conditions (3.19) and (5.1). Now, for any $0<\lambda<1$, the sequence $\bar{f}_\lambda^k \theta$, $k=1,2,\ldots$, converges uniformly to a function v^* in $\bar{V}$ having the following properties:

(i) $\bar{f} v^* = v^*$

(i') (equivalent to (i)) There is a constant g, called the gain or optimal performance, such that $[fv^*-v^*](\pi) = g$, $\forall\, \pi \in \Pi$

(ii) v^* is convex

(iii) $|v^*| \leq C$

(iv) $v^*(\pi) + K\,g - \max_{\pi' \in \Pi}\{v^*(\pi')\} \leq [f^K\theta](\pi) \leq v^*(\pi) + K\,g - \min_{\pi' \in \Pi}\{v^*(\pi')\}$

(v) $v^*(\pi) + g/(1-\beta) - \max_{\pi' \in \Pi}\{v^*(\pi')\} \leq \tilde{v}^*_\beta(\pi) \leq v^*(\pi) + g/(1-\beta) - \min_{\pi' \in \Pi}\{v^*(\pi')\}.$

Now (2.8), (2.9), (2.10) are immediate consequences of (4.10), (4.12), and (5.2).

VI. COMPUTATION OF AN ε-OPTIMAL CONTROLLER

Condition (3.19) implies that the state estimator can be arbitrarily closely approximated, for M sufficiently large, by a finite-state automaton that retains only the most recent M input-output pairs. (Compare this with a similar property of the stable Kalman filter).

ε-optimal finite-memory strategies may be computed by a method, called perceptive dynamic programming, based on this property of the state estimator. Suppose that the controller is able, at time k, to exactly measure the internal state at time k-M(k), and suppose moreover that the process M(k) is such that $\hat{z}(k) = [s(k-M(k)); (u(k-M(k)), y(k+1-M(k)), \ldots, (u(k-1), y(k))]$ is a sufficient statistic*. Then the problem can be expressed as an MDP having state process $\hat{z}(k)$; this is a simple generalization of [5]. Of course the resulting policy depends on information that is not available in practice, and so it cannot be considered a solution to the original problem. But the performance obtained is clearly an upper bound on feasible performance, since it assumes the availability of more information. Now the delayed state can be guessed and substituted into this policy. The resulting controller is feasible (when the guess $\hat{s}(k-M(k))$ is a function of the I/O pairs in $\hat{z}(k)$) and the closed loop system is now a Markov chain whose performance is readily evaluated. This performance is a lower bound on optimal feasible performance. It can be shown [11] that the difference between these bounds approaches zero as a lower bound on M(k) is increased. This algorithm will be discussed in detail in a later publication.

*For example, M(k) might be a constant. More generally, it suffices that M(k+1) be expressed as a (deterministic) function of $\hat{z}(k)$, u(k) and y(k+1) alone, and that $M(k+1) \leq M(k) + 1$.

VII. CONCLUSIONS

A finite-element plant model has been considered and controllability/observability-like conditions have been shown to imply well-posedness of the problem in the infinite-horizon case. A key concept in obtaining these results was a metric with respect to which the state estimator is a contraction. The author is currently interested in generalizing this metric to distributions on infinite state sets such as Euclidean space or the unit sphere. In the case of a Kalman filter, the contraction, in order to be analogous with what is presented here, must account not only for convergence of the conditional mean, but for convergence of the entire distribution to a normal distribution with appropriate covariance as well.

VIII. ACKNOWLEDGEMENT

I am grateful to Alvin W. Drake and to Sanjoy K. Mitter, my dissertation supervisors at MIT, for their continuing encouragement. This paper is based primarily on research conducted at the Decision and Control Sciences Group of the MIT Electronics Systems Laboratory, supported in part by AFOSR contracts 72-2273 and 77-3281, and reported in the author's doctoral dissertation.

REFERENCES

[1] Astrom, K. J., "Optimal Control of Markov Processes with Incomplete State Information," J. Math. Anal. Appl. 10, 1965.

[2] Astrom, K. J., "Optimal Control of Markov Processes with Incomplete State Information II: The Convexity of the Loss Function," J. Math. Anal. Appl. 26, 1969.

[3] Bar-Shalom, Y., and Tse, E., Dual Effect, Certainty Equivalence, and Separation in Stochastic Control, IEEE Trans. Aut. Contr. 19, 1974.

[4] Bertsekas, D., Dynamic Programming and Stochastic Control, Academic, New York, 1976.

[5] Brooks, D. M. and Leondes, C. T., "Markovian Decision Processes with State-Information Lag," Opns. Res. 21, 1973.

[6] Drake, A. W., "Observation of a Markov Process through a Noisy Channel," Sc.D. Thesis, Department of Electrical Engineering, M.I.T., Cambridge, MA, 1962.

[7] Feldbaum, A. A., Optimal Control Systems, Academic, New York, 1965.

[8] Isaacson, D. L. and Madsen, R. W., Markov Chains: Theory and Applications, Wiley, New York, 1976.

[9] Kushner, H. J., Introduction to Stochastic Control, Holt, Rinehart and Winston, New York, 1971.

[10] Paz, A., Introduction to Probabilistic Automata, Academic, New York, 1971.

[11] Platzman, L. K., "Finite Memory Estimation and Control of Finite Probabilistic Systems," Ph.D. Thesis, Department of Electrical Engineering and Computer Science, M.I.T . Cambridge, MA, 1977.

[12] Platzman, L. K., "Stability of State-Estimators for Partially-Observed Markov Chains," submitted to the IEEE Trans. Inf. Theory.

[13] Platzman, L. K., "Optimal Infinite-Horizon Undiscounted Control of Finite Probabilistic Systems," submitted to the SIAM Journal of Control and Optimization.

[14] Schweitzer, P. J., "Iterative Solution of the Functional Equations of Undiscounted Markov Renewal Programming," J. Math. Anal. Appl. 34, 1971.

[15] Smallwood, R. D. and Sondik, E. J., "The Optimal Control of Partially-Observable Markov Processes over a Finite Horizon, Opns. Res. 21, 1973.

[16] Sondik, E. J., "The Optimal Control of Partially-Observable Markov Processes over the Infinite Horizon: Discounted Costs," Opns. Res. 26, 1978.

[17] Sternby, J., "A Simple Dual Control Problem with an Analytical Solution," IEEE Trans. Aut. Contr. 59, 1976.

[18] Witsenhausen, H., "Separation of Estimation and Control for Discrete Time Systems," Proc. IEEE 59, 1968.

[19] White, D. J., "Dynamic Programming, Markov Chains, and the Method of Successive Approximations," J. Math. Anal. Appl. 6, 1963.